"By applying privacy threat modelling to ML/AI we have learned to humanize the machine. The combination of human and machine learning is clearly beneficial for the creation of safe, respectful and privacy friendly products." PLOT4ai

GUIDELINES

PLOT4ai is a library containing 86 threats classified under the following **8 categories**:

 Technique & Processes	**Our processes and/or technical actions can have a negative impact on individuals or cause harm**
 Accessibility	**We are not providing the ability to access and use our AI systems considering all type of individuals**
 Identifiability & Linkability	**Individuals can be linked to certain attributes or individuals and they can also be identified.**
 Security	**We can have a negative impact on individuals or cause harm by not protecting our AI systems and processes from security threats**
 Safety	**We do not recognize hazards and protect individuals from harms or other dangers**
 Unawareness	**We do not inform individuals and offer them the possibility to intervene**
 Ethics & Human Rights	**We do not reflect on matters of value and principles that can have a negative impact on individuals or cause harm**
 Non-compliance	**We do not comply with data protection law and other related regulations.**

PLOT4ai was created having LINDDUN GO as inspiration. It contains 5 new categories that are not part of LINDDUN: Technique & Processes, Ethics & Human Rights, Accessibility, Safety and Security.

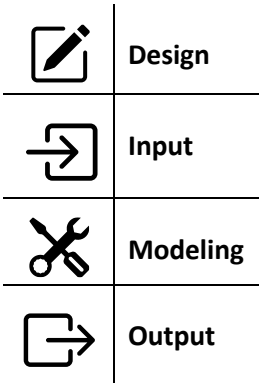
And it keeps some of the original categories from LINDDUN: Identifiability & Linkability, Unawareness and Non-compliance. As you can see, Identifiability & Linkability have been joined in

one category; the reason for this is that in my experience it is difficult for some people to distinguish between these two and in some scenarios, they can also be closely related.

Development Life Cycle:

In privacy it's essential to look at the data life cycle. ML/AI is an iterative process where data plays an important role in all the different phases and that is why it is important to check which development methodology you are using when creating models. This will also help you to capture more possible threats when threat modeling your AI system.

PLOT4ai contains a set of only **4 DLC phases**:



Comparing different methodologies (see table below), you can see that most of them consider similar phases that involve the whole data lifecycle.

When I created PLOT4ai I decided to create a set of simple phases to make it more accessible for non-technical stakeholders. And as you can see the building blocks are much simpler than the ones used in other methodologies, but they can be easily aligned.
For instance: The input phase would contain the *data understanding* and *data preparation* phases.

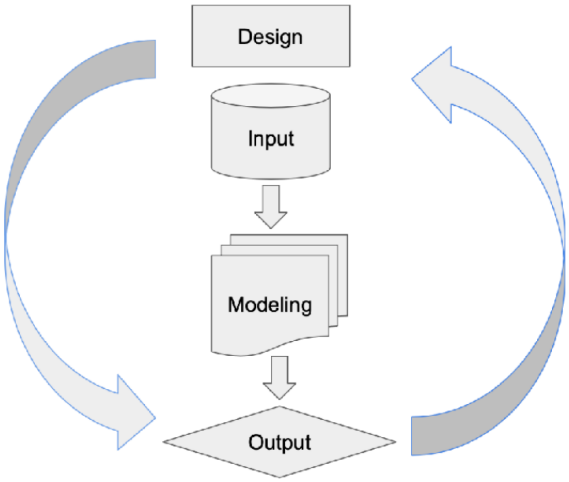
SEMMA	CRISP-DM	ASUM-DM	TDSP	MDM	PLOT4AI
	Business Understanding	Analyze	Business Understanding	Problem Mutation	Design
Sample Explore Modify	Data Understanding Data Preparation	Design	Data acquisition & understanding	Data Preparation Data understanding	Input
Model Assess	Modeling Evaluation	Configure & Build	Modeling	Model assembly Model audit	Modeling
	Deployment	Deploy Operate & Optimize	Deployment	Model delivery Maintaining & decommissioning	Output

SEMMA: Analytics Lifecycle, SAS Institute
CRISP-DM: Cross-industry standard process for data mining, ESPRIT
ASUM-DM: Analytics Solutions Unified Method, IBM
TDSP: Team Data Science Process, Microsoft
MDM: Model Development Process, Przemyslaw Biecek
PLOT4ai: Privacy Library Of Threats 😉

Data Flow Diagram

In threat modeling it is common practice to use a Data Flow Diagram (DFD) as a visual representation of the system you are going to analyse. A DFD offers you a good overview of the possible flows of data thereby making it easier to spot threats.

When threat modeling a ML/AI system you could start with a basic way of representing the data flow:



But you can also use a more complex representation, showing for instance the end points and interactions with other systems when you are collecting, receiving or sharing data, when you are storing your training cases and/or models in the cloud, when you are using ML libraries, when and where API's are used, etc…

For threats related to the categories: Technique & Processes, Identifiability & Linkability, Security and Non-Compliance, the use of a more detailed DFD is advisable. The DFD should be considered a live representation that will evolve as a result of your threat modeling analysis. Don't be afraid to start with a basic DFD, with the time it will hopefully transform into an actual representation of your AI systems and its interactions.

Threats as Cards

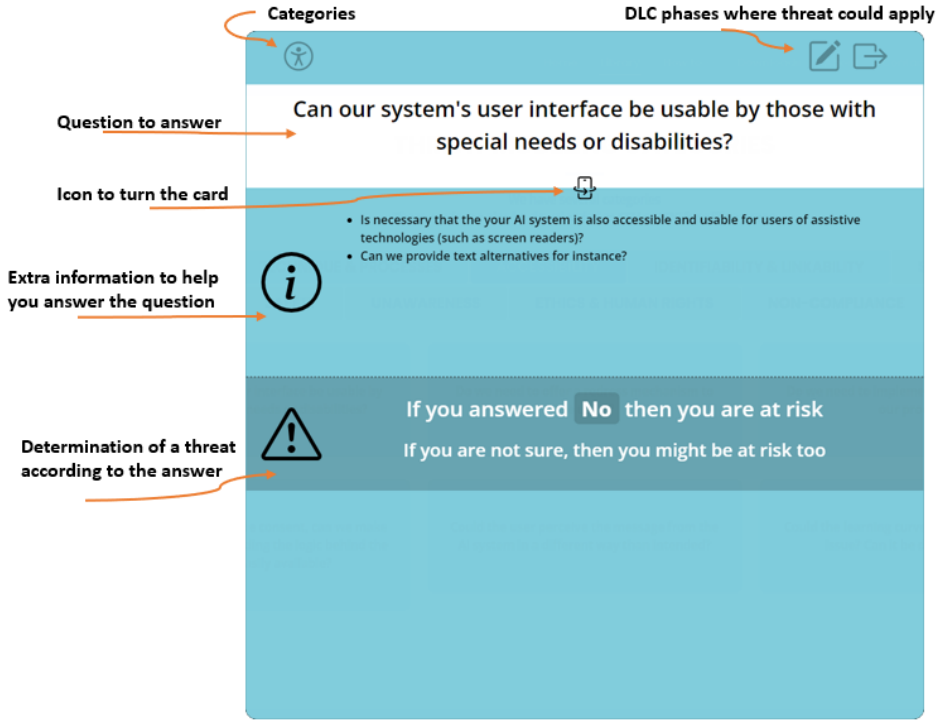
PLOT4ai provides threats in the form of cards just like LINDDUN GO does. The cards have different colours depending on the category they belong to.

Each card represents a possible threat brought to you as an elicitation question. During my observations I have seen better results when the threat is presented as a question compared to when it is presented as a description. This could be for several reasons, one of them being that some people find more difficult to visualise certain scenarios.

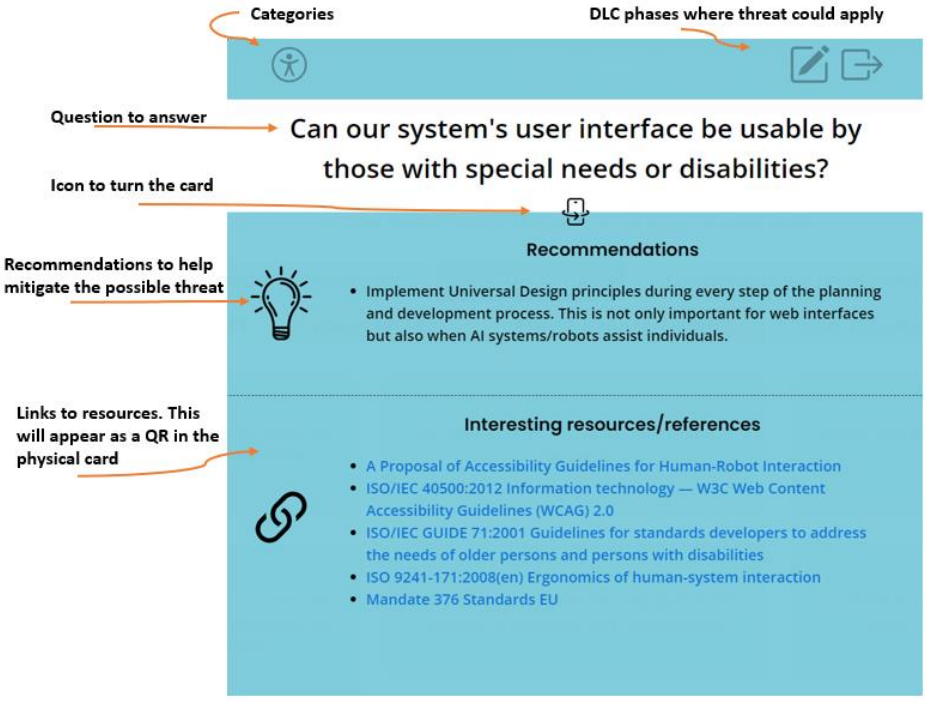
That is why I decided to take the question approach in PLOT4ai. **Questions** can help to bring the right focus and facilitate engagement. Especially in the design phase, where there is a lot to decide, this approach can be very beneficial.

This is what you can find on a card:

Card Front:



Card Back:



The cards have icons on the top representing the category they belong to and the Development Life Cycle where the threat could apply.

Why do some cards have more than one category icon?

All threats fall under a main category; this is represented by the colour of the card and the first icon on the left. But some threats could also fall under other categories; these categories appear as an extra icon on the right side besides the main category icon.

Why do some cards have more than one DLC icon?

Most of the threats can be applied to more than one phase of the Development Life Cycle.
Although my recommendation is to go through all the threats during the design phase, this might not be necessary if your organisation already has a strong quality process implemented and a certain organisational maturity level.
You can also find cards with all 4 icons of the DLC on the right side. The reason for this is that in my experience some threats have impact throughout the whole life cycle, and new events during the first phases could trigger changes that create new threats or open ones that were thought to be mitigated.

Card Deck:

The physical card deck contains 86 paper cards that are very similar to the digital version.
The only difference in the paper version is that all the cards contain a QR in the back that you can scan to see the online version. The links in the section "Interesting resources/references" have been replaced by the QR. The cards from the "Ethics & Human Rights" category containing a link to the EU Charter of fundamental rights also have a substitute QR code.

How can you apply PLOT4ai in practice?

PLOT4ai can be played like a card game using the physical card deck or the digital version of the cards. Simply choose what works best for you, relax and have some fun!

Quick tips before starting:

- Sessions should not be longer than 1.5, max. 2 hours to avoid tiredness and lack of focus. You can also do 30 min. timeboxed sessions focussing on just one specific category.
- It is important to identify all the relevant stakeholders that need to be present in the session. Especially during the design phase it is recommended that you involve all the people that have the knowledge and can take decisions. Remember that diversity is very important!
- A facilitator is needed to guide the sessions. Decide who will be taking this role. It does not have to be a privacy expert but having some knowledge can be helpful.
- Preparing for the session by selecting the right questions is very important. For instance, after the design phase, once the requirements are (more) clear, try to avoid selecting cards related to threats that are already taken care of during your quality assurance and control process. Not doing this will otherwise feel like a duplication of work and create frustrations.

- If prioritization of the threats is important, consider adding an extra column for Effort in the Threat Report Template (see step 8 below). The priority can then also take into account the effort that is required.
- This is like a game. Establish clear rules for time boxing: how long can discussions last per threat and when is an exception allowed.

Steps:

1. Gather a group of stakeholders to create a DFD of the system and interaction elements you want to analyse. A simple representation of the way the data might be flowing can be sufficient during the design phase. You could even jump into the threat modeling session without a DFD; depending on the use-case it is not always essential to have one.

2. Select cards for the session; you can randomly pick them or focus on a specific category. See also the Quick Tips.

3. With or without DFD, gather all the important stakeholders - now is when the actual threat modeling session will start.

4. For each selected card, read out loud the question and the extra info provided on the card.

5. Discuss the possible threat together. Time box how long you want to think about an answer: 2 minutes per answer can be sufficient but consider accepting exceptions if extra time is required because the group finds it difficult to reach consensus.
   When threat modeling the category Ethics & Human Rights consider giving more time per question. This category usually asks for a higher level of reflection in the group.

6. The card will indicate if answering YES or NO means that you have found a threat.
   If you are not sure, then it is always a possibility that you have found a threat.

7. If you have found a threat, turn the card to read the recommendations. This is optional though - you can also decide to do that after the session.

8. Document the threat. You can use the Threat Report Template that we provide for that.

9. Mark the question as a threat in the file and quickly discuss with the group if the threat should be classified as a Low, Medium or High risk. This is helpful to prioritize actions. You can also take the opportunity to write down some notes about possible actions and even indicate a (risk) owner.

	Risk				
Threat	Low	Medium	High	Actions	Owner
X		X			
X			X		

10. You are finished when time is over or when all cards are examined.

Next steps:

- Threats can also be added to your project backlog (in Jira por instance).
- You can decide to focus on easy/quick fixes first and later follow up on the rest.
- You will find threats that can be considered like a warning, but that are not really risks yet that you can mitigate at that moment. It is also important to document these threats and review them regularly.
- Consider establishing (privacy) acceptance criteria within your development team(s).
- In Agile: you can do privacy refinements to go through all the privacy user stories in the backlog.
- You can train your team in knowledge areas such as privacy, data protection and ethics. This can also be beneficial to facilitate the threat modeling sessions.

Benefits:

- Organisations can benefit from the fact that some of the threats play a more global role what will lead to a consequent improvement of processes. That is why it is important to register the threats and have an overview of what has been mitigated already. This can also be useful for KPI reporting.
- Another clear benefit is the reduction of rework: simply because the purpose is more clear and expectations regarding issues like bias, discrimination or explainability can be better managed.
- The threat modeling sessions also bring all stakeholders on the same page, reducing time spent in endless discussions.
- The output of the sessions can be also used in your (Data) Privacy Impact Assessments, what also saves time.
- It brings structure and focus to the teams, increases knowledge and collaboration.